# Condition of Lake Erie Largemouth Bass Sampled in the ODOW Nearshore Community Survey 2014-2017

*Alex Benecke*

*January 4, 2018*

### Data Prep

```
Stock <- read.csv("Data/Clean-Data/2012-2017_nearshore-survey-largemouth-bass_Stock_CLEAN.csv") %>%
  filter(fyr>=2014) %>%
  filter(!is.na(Wr)) %>%
  arrange(Year,gcat)

Stock$fyr <- as.factor(Stock$fyr)

headtail(Stock)
```

```
str(Stock)
```

```
## 'data.frame':    349 obs. of  16 variables:
##  $ Year  : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
##  $ Site  : int  8 10 18 15 16 6 10 15 16 2 ...
##  $ FID   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Weight: num  851 794 737 851 879 ...
##  $ Length: int  384 384 384 386 395 397 403 405 405 407 ...
##  $ AC    : int  3 3 3 3 3 3 3 3 3 3 ...
##  $ AGE   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ SexCon: int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Sex   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Delts : logi  NA NA NA NA NA NA ...
##  $ logW  : num  2.93 2.9 2.87 2.93 2.94 ...
##  $ logL  : num  2.58 2.58 2.58 2.59 2.6 ...
##  $ fyr   : Factor w/ 4 levels "2014","2015",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ Ws    : num  852 852 852 867 935 ...
##  $ Wr    : num  99.9 93.2 86.5 98.2 94 ...
##  $ gcat  : Factor w/ 3 levels "preferred","quality",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
unique(Stock$fyr)
```

```
## [1] 2014 2015 2016 2017
## Levels: 2014 2015 2016 2017
```

#### Note

*I removed the years 2012 and 2013. 2012 because only large fish have weight length data and more and differenct sites were samples. 2013 lacks weight data due to the loss of unifying ID variable and weight and lengths being seperated on different tabs.*

### Summarize Relative Weight by Year

```
(Wr.Stock <- Summarize(Wr ~ fyr, data = Stock) %>% arrange(fyr))
```

```
##    fyr   n     mean       sd   min     Q1 median    Q3   max
## 1 2014 140 109.6148 15.74476 80.40  98.78  106.9 117.8 151.3
## 2 2015  67 109.8385 15.56879 78.24  98.50  108.9 120.8 149.7
## 3 2016 107 115.3857 13.81954 61.76 107.90  115.4 124.6 146.2
## 4 2017  35 123.6056 33.66625 70.94 104.50  110.8 131.2 214.8
```

I have created a file with the relative weight of each gabelhouse length category for each year. The file name is relative-weight_largemouth-bass_STOCK.csv.

**Note**

*The relative weight data contains only stock length individuals. This is so that I can easily compare the relative weight of fish with PSD. This is done despite the min TL being 150 mm. I may want to summarize relative weight for 150mm and greater length individuals in the future to see if young/small fish drive down or increase Wr.*

Lets start exploring the relative weight data. I have two questions I would like to know the answer to.

1) does Wr differ among years?
2) does Wr differ among gabelhouse length categories?

**First** Lets see if Wr is different between years.

```
aov1 <- lm(Wr ~ fyr, data = Stock)
# save(aov1,file = 'model-output/aov1.rda')
Anova(aov1)
```

```
## Anova Table (Type II tests)
##
## Response: Wr
##           Sum Sq  Df F value    Pr(>F)
## fyr         6810   3  7.1691 0.0001108 ***
## Residuals 109235 345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mc1 <- glht(aov1, mcp(fyr = "Tukey"))
summary(mc1)
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = Wr ~ fyr, data = Stock)
##
## Linear Hypotheses:
##                  Estimate Std. Error t value Pr(>|t|)
## 2015 - 2014 == 0   0.2238     2.6434   0.085  0.99978
## 2016 - 2014 == 0   5.7709     2.2849   2.526  0.05561 .
## 2017 - 2014 == 0  13.9908     3.3627   4.161  < 0.001 ***
## 2016 - 2015 == 0   5.5471     2.7722   2.001  0.18418
## 2017 - 2015 == 0  13.7671     3.7111   3.710  0.00123 **
```

```
## 2017 - 2016 == 0   8.2199     3.4649   2.372  0.08144 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

Looks like Wr *is significantly different* between years (One-Way ANOVA, $F_{3,345} = 7.17$, $p < 0.001$). There is *no significant difference* in relative weight between 2015 and 2014 (Tukey HSD, t = 0.085, p = 1), 2016 and 2014 (Tukey HSD, t = 2.56, p = 0.06), 2016 and 2015 (Tukey HSD, t = 2.00, p = 0.18), and 2016 and 2017 (Tukey HSD, t = 2.37, p = 0.08). However, relative weight *is significantly different* between 2017 and 2014 (Tukey HSD, t = 4.16, p < 0.001), and 2017 and 2015 (Tukey HSD, t = 3.71, p = 0.001).

### constructing a plot of Wr and Year

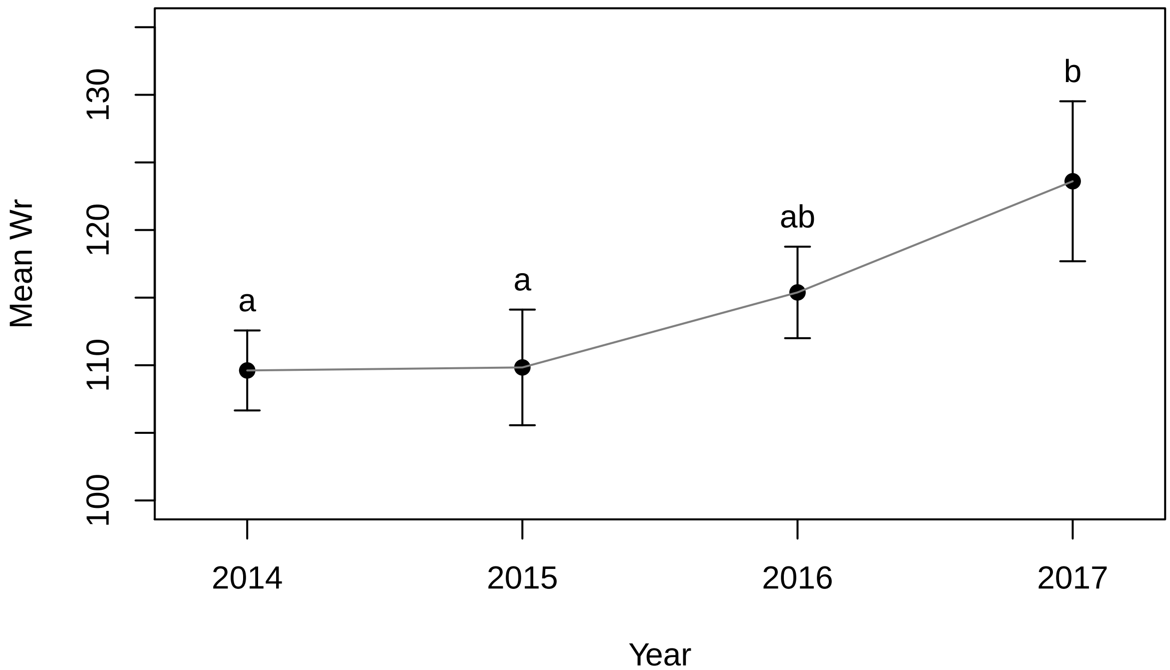
```r
grps <- c("2014", "2015", "2016", "2017")
nd <- data.frame(fyr = factor(grps, levels = grps))
(pred <- predict(aov1, nd, interval = "confidence"))
```

```
##        fit      lwr      upr
## 1 109.6148 106.6569 112.5727
## 2 109.8385 105.5628 114.1142
## 3 115.3857 112.0023 118.7691
## 4 123.6056 117.6898 129.5214
```

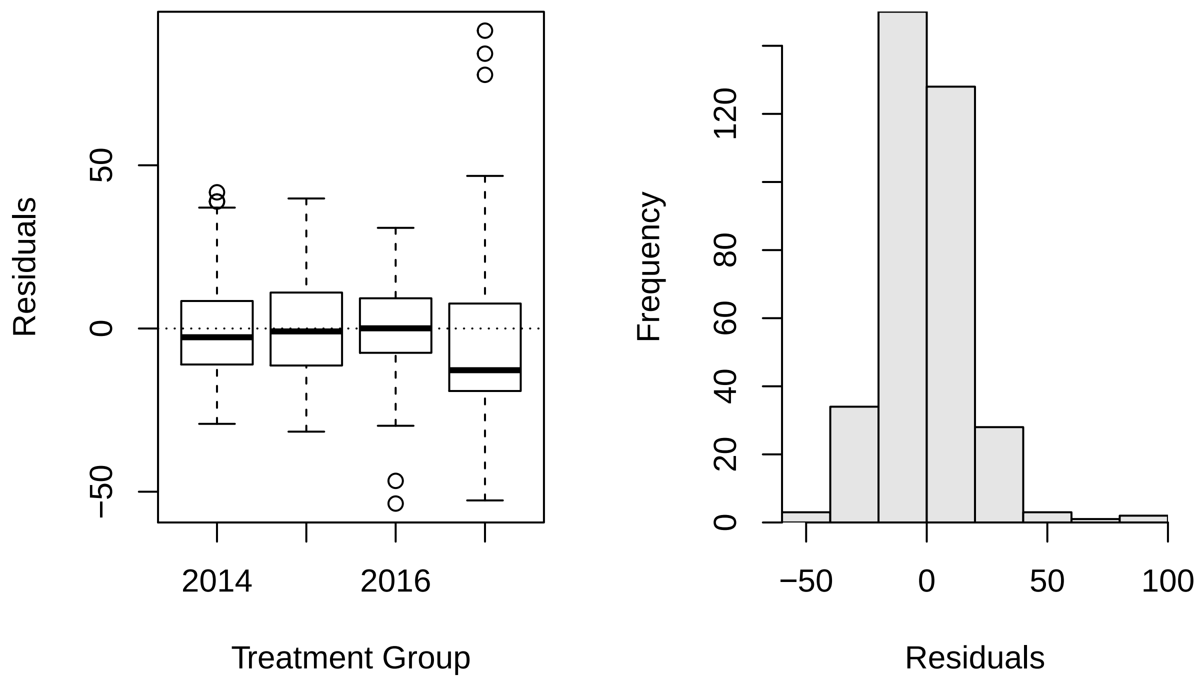
```r
plotCI(as.numeric(nd$fyr), pred[, "fit"], li = pred[, "lwr"], ui = pred[, "upr"],
    pch = 19, xaxt = "n", xlim = c(0.8, 4.2), ylim = c(100, 135), xlab = "Year",
    ylab = "Mean Wr")

lines(nd$fyr, pred[, "fit"], col = "gray50")

axis(1, at = nd$fyr, labels = nd$fyr)

cld(mc1)
```

```
## 2014 2015 2016 2017
##  "a"  "a" "ab"  "b"
```

```r
text(x = nd$fyr, y = pred[, "upr"], labels = c("a", "a", "ab", "b"), pos = 3)
```




```
residPlot(aov1)
```




```
leveneTest(aov1)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value    Pr(>F)
## group   3  9.9692 2.561e-06 ***
##       345
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This seems to suggest that the data violates the assumption that errors are normally distributed ($F_{3,345}$

= 9.97, p < 0.001).

**Without 2017** *Variance are equal and the homoscedasticity assumption is likley met (Levene's Test, $F_{2,311}$ = 1.75, p = 0.18).*

```
Year <- c("2014", "2015", "2016", "2017")
pred <- data.frame(Year, pred)
names(pred) <- c("Year", "Wr", "LCI", "UCI")
str(pred)
```

```
## 'data.frame':    4 obs. of  4 variables:
##  $ Year: Factor w/ 4 levels "2014","2015",..: 1 2 3 4
##  $ Wr  : num  110 110 115 124
##  $ LCI : num  107 106 112 118
##  $ UCI : num  113 114 119 130
```

```
head(pred)
```

```
##   Year       Wr      LCI      UCI
## 1 2014 109.6148 106.6569 112.5727
## 2 2015 109.8385 105.5628 114.1142
## 3 2016 115.3857 112.0023 118.7691
## 4 2017 123.6056 117.6898 129.5214
```

```
# 1-10-2018#write.csv(pred,file =
# 'Data/Clean-Data/relative-weight_largemouth-bass_STOCK.csv')
```

### Kruskal-Wallis Test

A non-parametric equivalent to a one-way ANOVA. This will test if distributions are similarly shaped with equal variances among groups, that the medians are equal among all groups.

```
kruskal.test(Wr~fyr,data = Stock)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Wr by fyr
## Kruskal-Wallis chi-squared = 17.519, df = 3, p-value = 0.0005527
```

This seems to sugest that the median for at least one group differs from the medians of one or more of the years (Kruskal-Wallis rank sum test, Kruskal-wallis chi-squared = 17.5, df = 3, p < 0.001).

```
dunnTest(Wr~fyr,data = Stock, method = "bonferroni")
```

```
## Dunn (1964) Kruskal-Wallis multiple comparison

##   p-values adjusted with the Bonferroni method.

##    Comparison          Z       P.unadj        P.adj
## 1 2014 - 2015 -0.4863958 0.6266865706 1.0000000000
## 2 2014 - 2016 -3.7703272 0.0001630337 0.0009782021
## 3 2015 - 2016 -2.6438110 0.0081978421 0.0491870528
## 4 2014 - 2017 -2.3869142 0.0169904581 0.1019427486
## 5 2015 - 2017 -1.8164101 0.0693074884 0.4158449307
## 6 2016 - 2017  0.1697604 0.8651985410 1.0000000000
```

```
dunnTest(Wr~fyr,data = Stock, method = "holm")
```

```
## Dunn (1964) Kruskal-Wallis multiple comparison

##   p-values adjusted with the Holm method.

##    Comparison          Z      P.unadj        P.adj
## 1 2014 - 2015 -0.4863958 0.6266865706 1.0000000000
## 2 2014 - 2016 -3.7703272 0.0001630337 0.0009782021
## 3 2015 - 2016 -2.6438110 0.0081978421 0.0409892107
## 4 2014 - 2017 -2.3869142 0.0169904581 0.0679618324
## 5 2015 - 2017 -1.8164101 0.0693074884 0.2079224653
## 6 2016 - 2017  0.1697604 0.8651985410 0.8651985410
```

**Holm adjusted p-value**

The Wr *is significantly different* between 2014 and 2016 (Dunn Multiple Comarison Test, Z = -3.77, $p < 0.001$), and 2015 and 2016 (Dunn Multiple Comarison Test, Z = -2.64, $p = 0.041$). However, Wr is not different between any other years.

### To Be Continued...

I will look into the difference in Wr between gcat at a later date. I don't think this matters so much as of now.

```r
Wr.14 <- filterD(Stock, Year == 2014)
Wr.15 <- filterD(Stock, Year == 2015)
Wr.16 <- filterD(Stock, Year == 2016)
```